California Seismic Safety Commission

Global Earthquake Model Foundation (GEM)

University of California at Los Angeles (UCLA)

"Back to Normal": Earthquake Recovery Modelling

November, 2016

Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

Recovery from the impacts of natural hazards, such as those caused by earthquakes, is a complex and multidimensional process, which constitutes the least understood component of the disaster management cycle. Not surprisingly, studies have shown that the time of recovery from earthquakes does not solely depend upon the extent of the initial damage, but it is strongly influenced by the socio-economic conditions of the affected region. To exemplify, low income families will likely depend on external financial aid (e.g. from government agencies) to start reconstruction, which often hinders the initiation of the rebuilding process. Moreover, residents with disabilities and the elderly will require special attention during the immediate response and long-term recovery phases following an event. These constitute some examples of how the social characteristics of a region determine the response and recovery phase following a damaging event.

The State of California, Alfred E. Alquist Seismic Safety Commission (CSSC) engaged the Global Earthquake Model Foundation (GEM) and the University of California at Los Angeles (UCLA), Department of Civil and Environmental Engineering to develop an operational framework and open source software that may be used for pre-earthquake planning and post-earthquake decision-making. The outcome is a set of methods and tools that are able to provide estimates of post-earthquake recovery times that consider physical damage from an earthquake as well as the effects of the socio-economic conditions on recovery. The city of Napa, California and the 2014 South Napa Earthquake were used as a real-world case study to demonstrate the tool and to validate the reliability of the results using a real-world recovery outcome.

1. Linking Damage and Pre-existing Socio-economic Conditions with Recovery Outcomes

While numerous research and disaster management communities have sought to explain long-term recovery processes, the ability to predict differential recovery outcomes is increasingly being seen as a key step for recovery planning and decision-making. To predict recovery outcomes, however, it is necessary to understand the determinants of recovery processes and how recovery from a damaging event, such as an earthquake, should be measured. To incorporate parameters that affect recovery outcomes from an earthquake into a quantitative framework and software tool, work was conducted to explore metrics of resilience and their association with differential recovery outcomes. Resilience is defined within this context as the ability of social systems to prepare for, respond to, and recover from damaging hazard events (Cutter, et al., 2008). It is within this context that a relationship between pre-existing socio-economic conditions (those associated with disaster resilience) and recovery is established through a novel framework, whereby disaster resilience indicators are used to predict the evolution of recovery (or recovery probabilities) of the building stock following an earthquake. The city of Napa, California and the 2014 South Napa Earthquake were used as a case study for the development and validation of the methodology. While relatively modest in intensity (M_w 6.0), this earthquake caused significant ground shaking and damage, particularly in the city's core. Direct costs associated with damage were estimated at USD 362 million, with economic costs to Napa County estimated at up to USD 1 billion (Galloway & Ingham, 2015). Figure 1-1 illustrates examples of affected residential properties, businesses and critical infrastructure damage.

Figure 1-1. Damage to residential properties, businesses and critical infrastructure. Photo credits: Napa Valley Register (J.L.Sousa), Justin Sullivan/Getty Images, KCRA (Brian Hicky), vos Iz Neias.

To define the relationship between the disaster resilience concept and earthquake recovery, a spatiotemporal assessment of recovery was statistically associated with metrics of disaster resilience that are cited in the research literature as being associated with recovery outcomes (Burton C. , 2015). The disaster resilience of the city of Napa is represented using a set of proxy variables that are classified into five subcomponents: social, economic, infrastructure, community capital and institutional resilience (Cutter, Burton, & Emrich, 2010). The variables, shown in Table 1-1, were retrieved from publically available sources at the census block group level of geography, as defined by the U.S. Census Bureau.

Table 1-1. List of variables representing pre-existing disaster resilience in the city of Napa.

Variables
Social Resilience
Percent of households where they speak English
Percent of housing units with no persons with a disability
Percent of the civilian noninstitutionalized population with any type of health insurance coverage
Percent of occupied housing units with telephone service
Percent of occupied housing units with vehicle available

Percent of the population 25 years and over that have at least a regular high school diploma
Percent of the total population that is male
Percent of the total population that is above 5 and below 60 years
Percent of the total population that is not a minority (White alone, not Hispanic or Latino)
Number of child care services
Percent of the total households with less than 5 persons
Percent of the single-parent households with a male householder, no wife present

Economic Resilience

Percent of households with earnings in the past 12 months
Percent of population 16 years and over in labour force that is employed
Percent of the population that has income in the past 12 months at or above poverty level
Per capita income as a fraction of the highest amongst the block groups
Percent of the renter-occupied housing units with gross rent less than $1500[a)] (+)
Percent of the civilian employed population 16 years and over that are not employed in food, accommodation and retail trade[b)]
Percent of females 20 to 64 years in households that are in labour force
Percent of occupied housing units that are owner occupied
Percent of the civilian employed population 16 years and over that are employed in healthcare practitioners and technical occupations
Percent of households with no supplemental security income in the past 12 months
Percent of households with no public assistance income in the past 12 months

Infrastructure Resilience

Housing density
Percent of housing units that are built after 1950
Percent of housing units that are not mobile homes
Number of internet, television, radio and telecommunications broadcasters
Number of schools (primary and secondary)
Number of hotels & motels
Number of banks
Percent of housing units that are vacant
Number of police, fire, emergency relief services and temporary shelters

[a)] In the U.S., it is commonly accepted that families who pay more than 30% of their income for housing are considered cost burdened. The value of $1500 as a limit of affordability was set according to this rule.
[b)] This variable is used as a proxy for single sector employment dependence.

Percent of the housing units that are single family detached homes
Community Capital
Number of civic and social advocacy organizations
Number of churches and religious organizations
Number of arts, entertainment and recreation centres, libraries, museums, parks and historic sites
Percent of the population that lived in the same Metropolitan Statistical Area 1 year ago
Institutional Resilience
Percent of the civilian employed population 16 years and over employed in emergency services (firefighting, law enforcement, protection)

The spatiotemporal evaluation of the recovery in the city of Napa was accomplished using *in situ* observations of building damage at six-month intervals. In the aftermath of the 2014 Napa Earthquake, building damage observations (i.e. location, building type, color-tagging information and damage description) were geocoded by city officials and made available via a web Geographic Information System (GIS). This information was retrieved to build a geospatial point-level dataset of 1462 damaged buildings. After the development of the initial damage database, three separate field surveys were conducted in the city of Napa, 6 months, 12 months and 18 months following the earthquake. As part of the recovery evaluation process, a detailed inspection of a set of 356 damaged buildings was conducted for which different recovery stages were attributed on a building-by-building basis. Due to time constraints, it was not feasible to survey all the damaged buildings (1462); therefore, 356 were selected for inspection (Figure 1-2). These included all the red-tagged and a random sample of the yellow tagged structures.

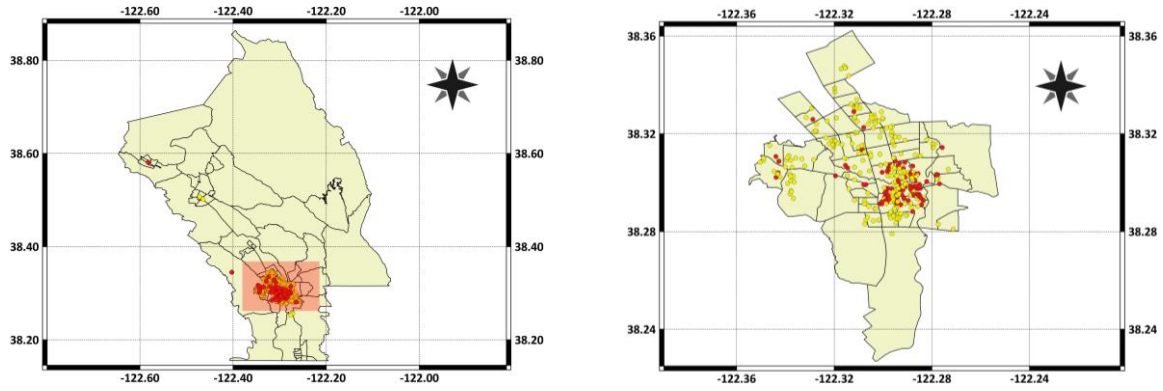


Figure 1-2. (a) Initial red and yellow tagged buildings across the block groups of the County of Napa and (b) the 356 evaluated red and yellow tagged buildings (zoom in the city of Napa).

At each field evaluation, each building was assigned a binary code (0 or 1) defining the stage of

the recovery, based on an exterior visual inspection. In this framework, 0 represents a "No Recovery" stage; and 1 is associated with the building's "Full Recovery" in which the building is fully repaired/rebuilt and reoccupied (Figure 1-3). Recovery from an earthquake depends on several factors and could include social capacities, the financial reserves of individuals and communities, political decision-making, the severity of damages sustained, and the proportion of a community adversely affected. The validation metric for this case study focuses explicitly on the material manifestation of recovery (i.e., the reconstruction of residential and commercial buildings), although this work is sensitive to the multifaceted nature of recovery. The rationale for considering the reconstruction of the built environment is that reconstruction is essential for returning life and livelihoods in Napa to pre-impact levels of functioning.

(a) Recovery Category "0" **(b)** Recovery Category "0" **(c)** Recovery Category "1"

Figure 1-3. Example of the recovery progress of one building in the city of Napa after: (a) 6 months; (b) 12 months and (c) 18 months following the earthquake.

According to the first inspection (six months following the event), almost 61% of the initial yellow-tagged buildings and nearly a quarter of the initial red-tagged structures were fully repaired. One year after the earthquake, the structures previously classified as fully recovered were excluded from the evaluation; the remaining buildings were re-evaluated to define an updated recovery category, based on their reconstruction progress. During the second field survey, the recovery stage of thirty-six buildings evolved from category "0" to "1", with equal proportion between yellow and red tagged structures. Finally, during the third survey, 18 months following the earthquake, fifty-five additional buildings were fully recovered, thirty-nine of which were initially yellow tagged and sixteen red tagged.

To determine the relationship between the collected set of disaster resilience metrics and the observed recovery outcomes over time, a parametric probabilistic model was proposed, allowing the treatment of uncertainties in a robust and statistically significant way. Specifically, a logistic regression model (Hosmer & Lemeshow, 2000) was calibrated to predict the probability of a "Full

Recovery" occurring in each of the block groups for which resilience variables were collected and recovery outcomes were observed. The socio-economic parameters constitute the independent variables (predictors) of the regression, while the recovery observations are the dependent (response) variables. Because the temporal evolution of the recovery process is also of interest, the variable time was included as a predictor, assuming values of 6, 12 and 18 months. Finally, these parameters were complemented with an indicator able to implicitly account for the level of seismic-induced damage at a given location. Specifically, the median Modified Mercalli Intensity (MMI) (Wood & Newmann, 1931) observed in each block group at the time of the event was utilized as a measure of seismic-induced damage. The distribution of MMI is readily available and was acquired from the United States Geological Survey (USGS) ShakeMap platform at a spatial resolution of 30 arc seconds.

For the regression to be possible, each block group was assigned a recovery stage (0 or 1), to make the data analogous to the resilience variables that were collected at the block group level of geography. To accomplish this, a simulation procedure was devised, resulting in a distribution of recovery probabilities for each block group, for each time, as a function of the selected set of independent variables. In other words, not only an absolute value, but a distribution of recovery probabilities in each block group is computed, capturing the uncertainty associated with the random nature of the recovery process (Figure 1-4).

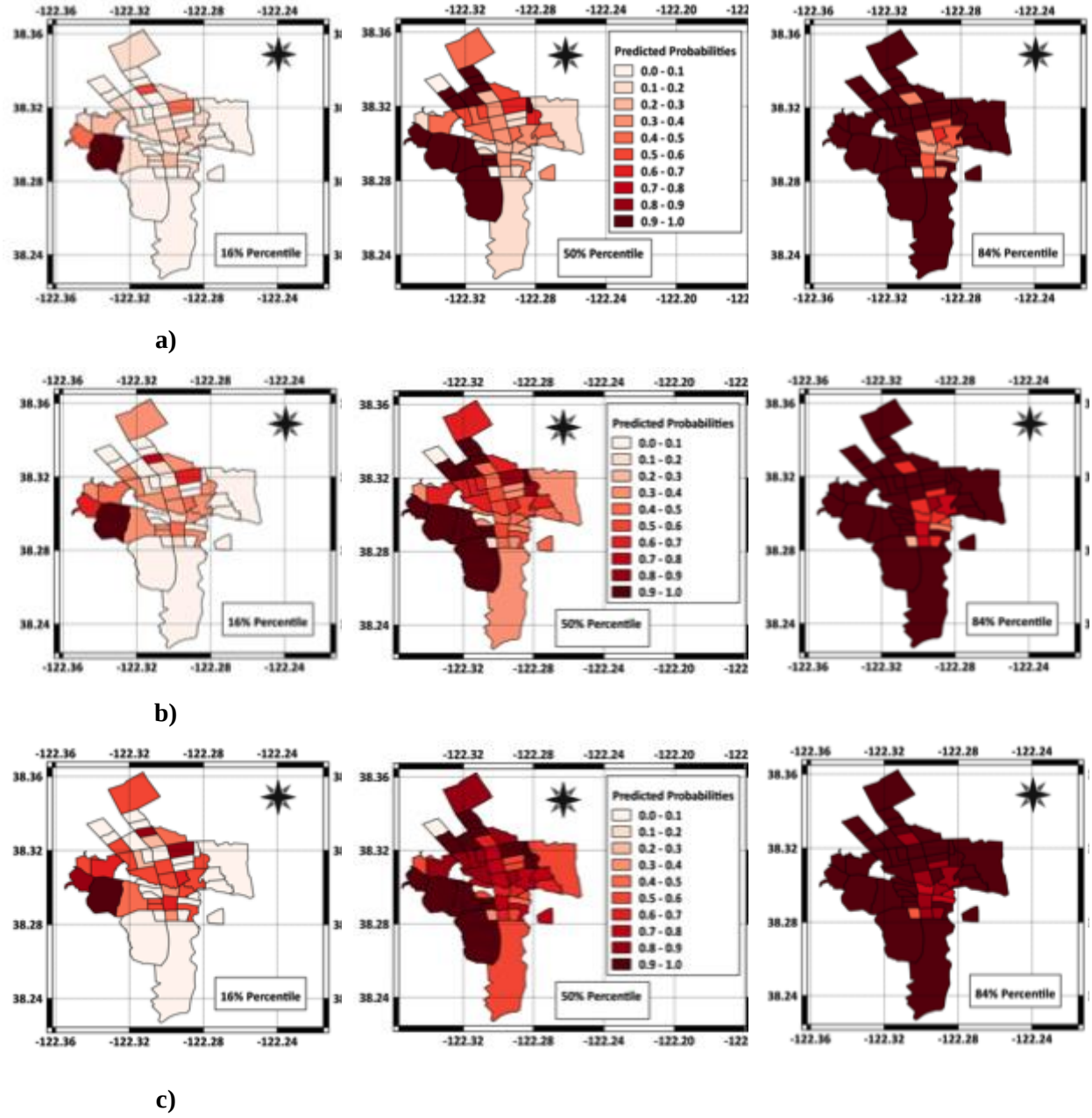


Figure 1-4. 50% (median), lower (16% percentile) and upper (84% percentile) bound recovery probabilities in the city of Napa as determined by the model, at a) 6 months, b) 12 months, and c) 18 months following the earthquake.

As a measure of the quality of the results, the mean predicted values of recovery are plotted as a function of the "surveyed" probabilities in each block group for each of the field surveys (Figure 1-5). In this context, "surveyed" probabilities are determined as the fraction of assessed buildings that are fully recovered at the time of interest, in each block group. As illustrated in Figure 1-5, the relationship between mean predicted and surveyed values approaches an almost perfect linear trend for the three survey instances.

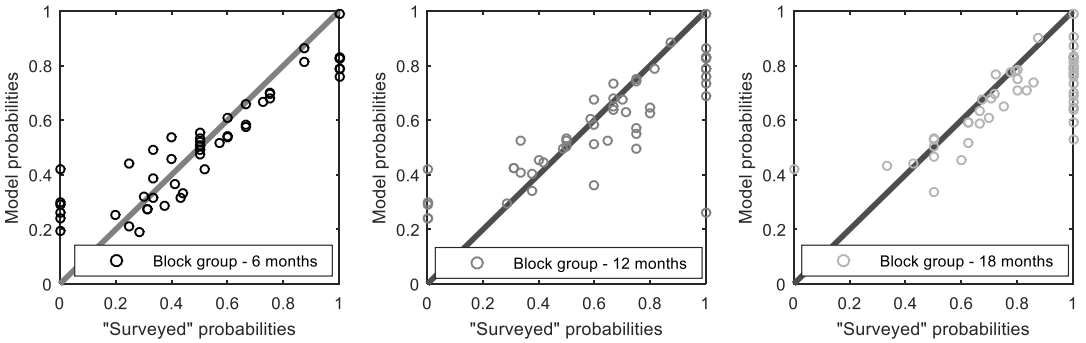


Figure 1-5. Predicted versus "surveyed" probabilities in each block group; for the first, second, and third field surveys.

The methodology further allows us to identify the resilience variables that most strongly affect the recovery trajectory, as depicted by our model (Table 1-2).

Table 1-2. Independent resilience variables contributing more strongly to the predicted recovery probabilities.

Variables	Acronym
Social Resilience	
Percent of households where they speak English	PH-ENGLISH
Percent of the civilian noninstitutionalized population with any type of health insurance coverage	PHEALTHINS
Percent of the single-parent households with a male householder, no wife present	PMALE-HOUSEHOLDER
Economic Resilience	
Percent of households with earnings in the past 12 months	PH-EARNINGS
Percent of population 16 years and over in labour force that is employed	P-EMPLOYED
Percent of occupied housing units that are owner occupied	PHU-OWNED
Infrastructure Resilience	
Percent of housing units that are built after 1950	PHU-BUILTAFT50
Post-earthquake damage and Time	
Modified Mercalli Intensity	MMI
Values of 6, 12, 18 months	TIME

According to Table 1-2, seven of the thirty-eight resilience parameters, the variable representing the time, and the damage indicator (MMI) significantly contribute to the prediction of the recovery trajectory occurring in Napa. These resilience parameters correspond to the social, economic and infrastructure resilience subcomponent.

To better understand the contribution of the most "significant" resilience variables, Figure 1-6 illustrates the spatial distribution of the six most significant indicators. In general, lower values are identified in the central part of the city for all the six indicators, when compared with those of the outer block groups. This pattern is in accordance with the median predicted recovery outcome, according to which lower recovery probabilities are also found in the central block groups (Figure 1-4). Conversely, the distribution of MMI has an opposite effect on the recovery process. More specifically, MMI values appear to be higher and more homogeneous in the central block groups, where a slower recovery process is taking place. Not surprisingly, less resilient and more damaged areas are linked to a longer recovery process.

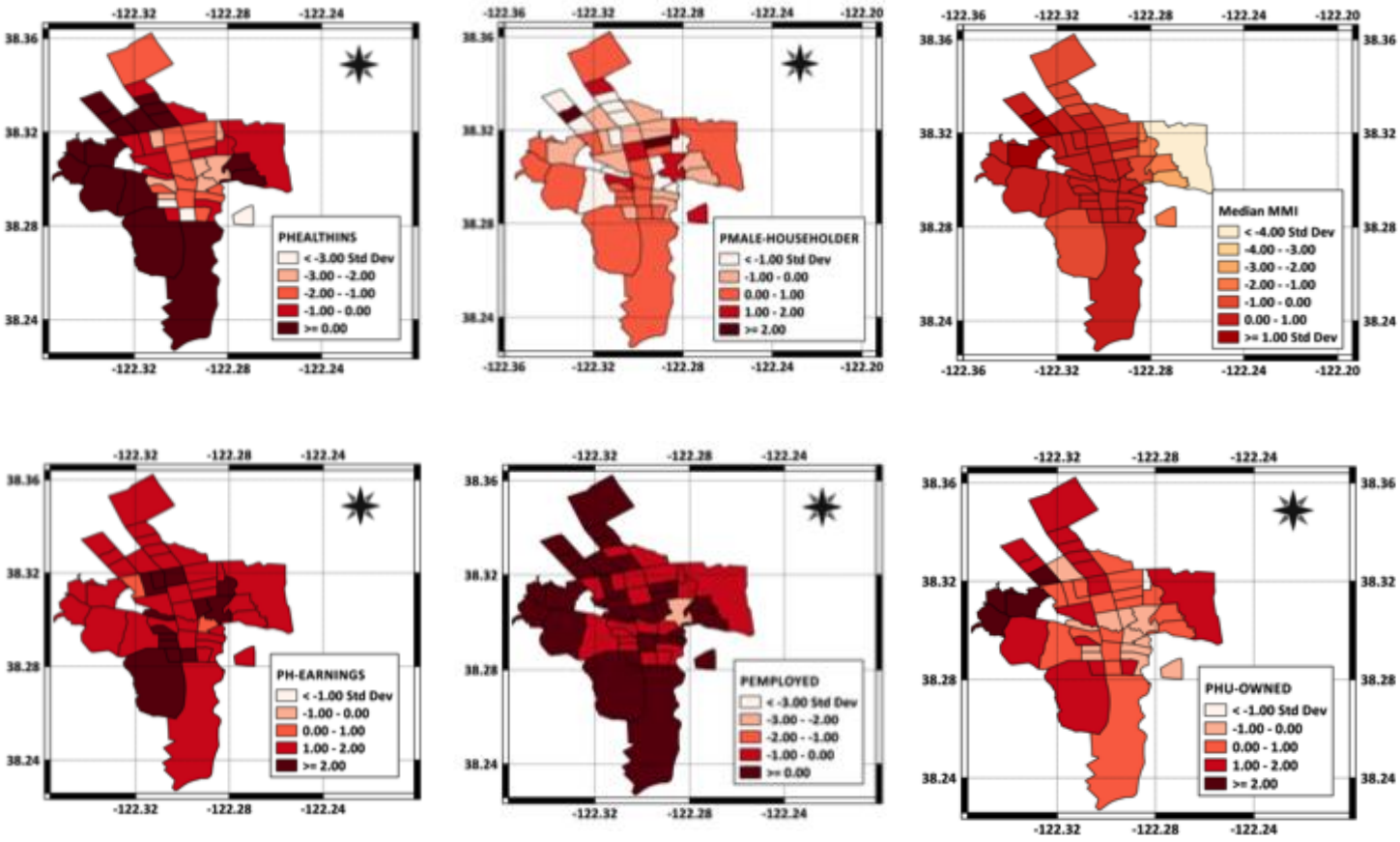


Figure 1-6. Spatial distribution of six of the nine significant variables presented in Table 1-2.

Practicality of the case study

As previously described, seven of the thirty-eight selected resilience indicators, the variable time, and the spatial distribution of MMI were verified to be the parameters that contribute to the prediction of building recovery based on their statistical significance. More specifically, the seven

identified resilience variables are: the percentage of **English speaking** population, the presence of **health insurance** coverage, the **households' structure**, the households' **earnings** (whether a household has any type of available income), **employment status**, **homeownership**, and the percentage of **buildings constructed after 1950**.

One of the main benefits of the proposed methodology is its potential to be applied in different areas affected by an earthquake for which damage/recovery data is not available. This is possible since the model allows the prediction of the recovery over time, solely by providing the selected set of resilience variables and the spatial distribution of the MMI. This overcomes the problem arising from the usual lack of post-earthquake damage and/or recovery observations over time and provides decision-makers and stakeholders an immediate overview of the recovery progress of their community. Thus, decision-makers can identify vulnerable areas that lag to recover and define their actions accordingly (e.g. anticipate the need of temporary sheltering).

In addition, the indication of the most "significant" socio-economic drivers of the recovery, along with the spatial distribution and evolution of the predicted recovery probabilities, facilitates the identification of socio-economic weaknesses and strengths within communities. This allows key stakeholders and policy-makers to identify a priori which are the areas that, due to their increased vulnerability, are more prone to experience greater difficulties in recovering. Thus, such information can be utilised to develop pre-disaster recovery plans that reflect the actual needs of the population and contribute to a rapid and efficient transition to normality. In addition, following an earthquake, the proposed methodology can be used by stakeholders to re-define or enforce rehabilitation efforts, by comparing the recovery predictions with the actual recovery trajectory. To exemplify, an area that is recovering considerably slower than predicted may imply that additional support to the vulnerable groups and better management in the indicated regions is required.

2. Post-Earthquake Recovery Modelling Methodology

Overview

A major earthquake occurring in one of the many large urban centers of California could lead to thousands of casualties, hundreds of thousands of displaced households and billions of dollars in losses. The lives of the impacted residents are likely to be enormously disrupted. The pace of recovery will depend among other things on the extent of building and lifeline damage, the extent of business disruption, the availability of utilities and how quickly communities can repair and replace their housing. Recent disasters like Hurricane Katrina and Super Storm Sandy have demonstrated the need to facilitate the speedy recovery of permanent housing in the affected communities. The immediate impact and pace of housing recovery is directly related to the likelihood of permanent outmigration of residents from the region. The overall goal of this section is to demonstrate the scientific framework and computational tools developed to quantify the effectiveness of specific resilience-building actions (preparedness, mitigation, and response) that would increase the speed of recovery following an earthquake.

The four main components of the recovery modelling methodology are (1) recovery-based limit state fragility function development, (2) developing building-level time dependent recovery functions, (3) accounting for the effect of externalities and socio-economic vulnerability and (4) developing community/regional level recovery functions. These are discussed in detail in the following subsections.

Recovery-Based Fragility Function Parameters

A rigorous evaluation of seismic resilience requires probabilistic methods for assessing limit states that influence post-earthquake functionality, which can be incorporated in modelling the recovery of the building stock. The methodology incorporates a set of building performance limit states that specifically inform community seismic resilience (Figure 2-1). These limit states have been adapted from the building performance categories defined by SPUR. They include (i) damage triggering inspection, (ii) occupiable damage with loss of functionality, (iii) unoccupiable damage, (iv) irreparable damage and (v) collapse. These limit states are different from those that are currently used in OpenQuake and other risk modelling platforms. This sub-section is intended to document the methodology used to map the fragility function parameters from the loss-based limit states used in OpenQuake to those of the recovery-based limit states used to model recovery.

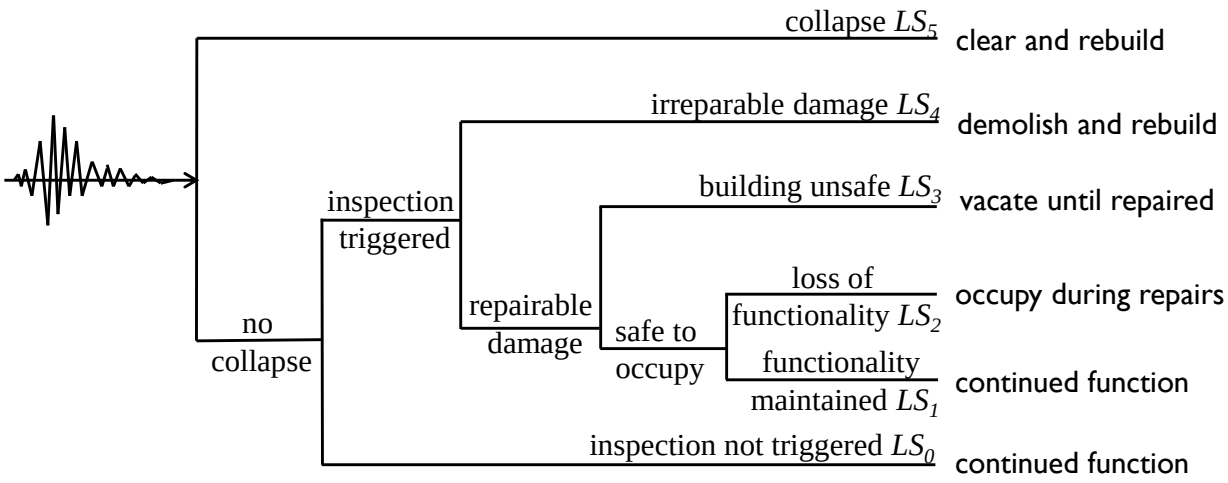


Figure 2-1. Event tree showing building performance limit states and recovery actions.

Recovery-Based Building Performance Limit States

Five discrete limit states (*LS0 through LS5*) are used, which are explicitly linked to post-earthquake recovery-related activities. Each limit state is associated with a unique combination of the following consequent actions to restore building function:

- Assessment and planning activities i.e. post-earthquake inspection and/or evaluation, preparation of plans and designs, financing and bidding preparation for construction work;
- Repairs needed to make building occupiable and repairs needed to restore functionality for repairable buildings; and
- Demolition and building replacement for non-repairable buildings.

LS0 - Damage below the threshold that would trigger inspection.

LS1 - Damage Triggering Inspection with Functionality Maintained: This represents the minimum damage threshold that would require post-earthquake inspection and/or evaluation. It is also used to imply a level of damage where, despite the need for post-earthquake inspection, the structural safety and critical subsystems essential to the functionality of the building are not compromised. However, operations may be impacted if the owner/operator decides to close the facility until inspections are completed. This decision is prompted by visible damage to structural (cracking of concrete members) or non-structural elements (partitions, facades, etc.). This type of damage occurs at low drift levels and affects structural and non-structural components with low deformation capacities.

LS2 - Occupiable Damage with Loss of Functionality: This implies that the building is structurally safe, occupiable and accessible but unable to carry out its primary function. This loss of functionality can occur despite the preservation of structural integrity as a result of damage to building systems, non-structural components or contents, which are critical to the operations of the facility. There may also be damage to structural components whose repair actions hinder normal building operations.

LS3 - Unoccupiable Damage: This infers that the building is either inaccessible or not safe to occupy following an earthquake. The loss of structural safety will likely be due to a substantial loss in the load carrying capacity of the gravity or lateral system that poses a life safety threat in the event of an aftershock. It is also possible but less likely for non-structural damage to compromise the safety or prevent access to the building. This is usually in the form of some type of falling hazard (e.g. brick façade or infill panels); however, these types of dangers can be mitigated in a short period of time. *LS3* is of particular importance to residential buildings as it is directly related to the shelter-in-place performance goal, emphasized in SPUR's resilient city initiative.

LS4 - Irreparable Damage: LS4 pertains to cases where the building is damaged to such an extent that repair becomes technically or cost prohibitive, necessitating demolition and replacement. The three main earthquake-related situations that can lead to demolition include (1) large permanent deformations and story drifts that make repairs unfeasible, (2) direct economic losses that exceed the limit set by insurance providers triggering full-value pay-out leading to complete replacement and (3) damage to key structural components that could significantly impede the repair process.

LS5 - Collapse: LS5 is related to complete or partial collapse, which is generally associated with either excessive lateral deformations (sidesway collapse) or the local or global loss of vertical load carrying capacity.

Loss-Based Building Performance Limit States

Risk modelling platforms, such as OpenQuake and HAZUS, use limit state fragility functions that relate earthquake ground shaking intensity to building damage. These limit states are used to link ground motion intensity to direct economic losses (vulnerability curves) that result from having to repair or replace damaged buildings. The limit states are classified based on construction type and are described in terms of the type and extent of physical damage to the building. The following is a description of the limit states (taken from HAZUS), which are relevant to wood frame single- and multi-family residential buildings found in Napa and California in general:

Slight Damage: Small cracks in non-structural elements (window, wall intersections, masonry chimneys, masonry veneer and stucco) and slippage in bolted connections.

Moderate Damage: Small cracks across shear wall, large cracks at doors, windows and masonry veneer, topping of tall masonry chimneys, minor slack in diagonal rod bracing and small cracks and split in bolted connections.

Extensive Damage: Large cracks across shear wall plywood joints, large slack at diagonal and broken braces, permanent lateral movement at floors and roof, topping of most brick chimneys, small cracks in foundations, split and/or slippage of sill plates and partial collapse at garage with soft-story configurations.

Complete Damage: Large permanent lateral displacement, may collapse, imminent collapse, some structures slip off foundations, large foundation cracks, 3% total area collapsed, broken brace rod or failed framing connections.

Methodology for Mapping Fragility Function Parameters from Loss-Based to Recovery-Based Building Performance Limit States

There is an obvious correlation between the loss-based and recovery-based building performance limit states. In both cases, the limit states are discrete, sequential and mutually exclusive with the higher limit states being associated with more extensive damage. This obvious link was used as the basis for mapping the fragility function parameters between the two types of limit states.

The fragility function for each of the loss-based limit states is assumed to take on a lognormal distribution and is defined by the following relationship:

$$P(DS > ds_i \mid S_d) = \Phi\left[\frac{1}{\beta_{ds_i}} \ln\left(\frac{S_d}{\bar{S}_{d,ds_i}}\right)\right] \quad \textbf{(2-1)}$$

where $\bar{S}_{d,ds_i}$ is the median value of the spectral displacement at which the building reaches the threshold of damage state ds_i, β_{ds_i} is the standard deviation of the natural logarithm at which the building reaches the threshold of the damage state ds_i and Φ is the standard normal cumulative distribution function. For a given building construction type, HAZUS provides the median spectral displacement and log standard deviation for each of the four loss-based limit states. The interstory drift at the threshold of each limit state is also provided. In addition to the building construction

type, the parameters also vary based on the seismic code design level which is directly related to the age of the buildings. Four code design levels are included, high-code, moderate code, low code and pre-code. The fragility function parameters for the two construction types for wood frame buildings (W1, wood light frame and W2 wood commercial and industrial) are shown in Table 2-1 (taken directly from HAZUS). The parameters are also separated based on seismic code design level.

Table 2-1. Fragility function parameters for loss-based damage states for wood frame buildings: (a) High-Code, (b) Moderate Code, (c) Low-Code and (d) Pre-Code seismic design levels.

Building Properties			Interstory Drift at Threshold of Damage State				Spectral Displacement (inches)							
Type	Height (inches)						Slight		Moderate		Extensive		Complete	
	Roof	Modal	Slight	Moderate	Extensive	Complete	Median	Beta	Median	Beta	Median	Beta	Median	Beta
W1	168	126	0.0040	0.0120	0.0400	0.1000	0.50	0.80	1.51	0.81	5.04	0.85	12.60	0.97
W2	288	216	0.0040	0.0120	0.0400	0.1000	0.86	0.81	2.59	0.88	8.64	0.90	21.60	0.83

(a)

Building Properties			Interstory Drift at Threshold of Damage State				Spectral Displacement (inches)							
Type	Height (inches)						Slight		Moderate		Extensive		Complete	
	Roof	Modal	Slight	Moderate	Extensive	Complete	Median	Beta	Median	Beta	Median	Beta	Median	Beta
W1	168	126	0.0040	0.0099	0.0306	0.0750	0.50	0.84	1.25	0.86	3.86	0.89	9.45	1.04
W2	288	216	0.0040	0.0099	0.0306	0.0750	0.86	0.89	2.14	0.95	6.62	0.95	16.20	0.92

(b)

Building Properties			Interstory Drift at Threshold of Damage State				Spectral Displacement (inches)							
Type	Height (inches)						Slight		Moderate		Extensive		Complete	
	Roof	Modal	Slight	Moderate	Extensive	Complete	Median	Beta	Median	Beta	Median	Beta	Median	Beta
W1	168	126	0.0040	0.0099	0.0306	0.0750	0.50	0.93	1.25	0.98	3.86	1.02	9.45	0.99
W2	288	216	0.0040	0.0099	0.0306	0.0750	0.86	0.97	2.14	0.90	6.62	0.89	16.20	0.99

(c)

Building Properties			Interstory Drift at Threshold of Damage State				Spectral Displacement (inches)							
Type	Height (inches)						Slight		Moderate		Extensive		Complete	
	Roof	Modal	Slight	Moderate	Extensive	Complete	Median	Beta	Median	Beta	Median	Beta	Median	Beta
W1	168	126	0.0032	0.0079	0.0245	0.0600	0.40	1.01	1.00	1.05	3.09	1.07	7.56	1.06
W2	288	216	0.0032	0.0079	0.0245	0.0600	0.69	1.04	1.71	0.97	5.29	0.90	12.96	0.99

(d)

The primary objective of this part of the overall study was to map the loss-based fragility parameters shown in Table 2-1 to recovery-based limit state parameters. The first step was to estimate the conditional probability of being in a particular recovery-based limit state given the occurrence of a loss-based damage state. This conditional probability relationship is defined as follows:

$$P(RBDS = rbds_i \mid LBDS = lbds_j) \quad \textbf{(2-2)}$$

where $P(RBDS > rbds_i \mid LBDS_j = lbds_j)$ is the probability that the recovery-based damage state $(RBDS)$ i occurs given that the loss-based damage state $(LBDS)$ j has been observed. Estimates of these conditional probabilities are provided in Table 2-2. The current values are based on

engineering judgement. They were obtained by examining the physical description of damage provided for the loss-based limit states and inferring the likelihood that this type of damage would trigger each of the six (*LS0* through *LS5*) recovery-based limit states. Later on in the development of the framework, these estimates were refined based on the results from nonlinear response history analyses of typical wood frame buildings using the *OpenSees* modelling platform. The results from the structural response simulation were used to establish analytical fragility functions for both types of limit states. This process enabled the development of a more explicit relationship between the fragility parameters for the two types of limit states. The conditional probabilities estimates can be further refined using heuristic data obtained from expert opinion. However, this approach is outside the scope of the current project.

In Table 2-2, each row provides the probability of being in each of the recovery-based limit states given the occurrence of the loss-based limit state in the first column of that row. For example, it can be observed that for a building that is in the loss-based limit state corresponding to moderate damage, the probability of being in recovery-based limit states *LS0, LS1, LS2* and *LS3*, is 0.2, 0.4, 0.3 and 0.1 respectively with a zero probability of being in the remaining limit states (*LS4* and *LS5).* Given that the recovery-based limit states are mutually exclusive and collectively exhaustive, each row must sum to one.

Table 2-2. Conditional probabilities used to map fragility parameters for loss-based to recovery-based limit states.

Loss-Based Damage States	$P(RBDS = rbds_i \mid LBDS = lbds_j)$					
	LS0 Inspection not Triggered	*LS1* Inspection	*LS2* Loss of Functionality	*LS3* Unsafe to Occupy	*LS4* Damaged Beyond Repair	*LS5* Collapse
None	1.0	0.0	0.0	0.0	0.0	0.0
Slight	0.6	0.4	0.0	0.0	0.0	0.0
Moderate	0.2	0.4	0.3	0.1	0.0	0.0
Extensive	0.0	0.0	0.2	0.4	0.3	0.1
Complete	0.0	0.0	0.0	0.0	0.2	0.8

Given the loss-based fragility function parameters in Table 2-1 and the conditional probability estimates in Table 2-2, the probability of occurrence of a particular recovery-based limit state can be obtained using the total probability theorem:

$$P(RBDS = rbds_i \mid S_d) = \sum_{j=1}^{n_{lbds}} P\left(PRBDS = rbds_i \mid LBDS = lrbds_j\right) \cdot P\left(LBDS = lrbds_j \mid S_d\right) \quad \textbf{(2-3)}$$

where $P\left(PRBDS = rbds_i \mid LBDS = lrbds_j\right)$ is taken from Table 2-2 and $P\left(LBDS = lrbds_j \mid S_d\right)$ is obtained from the fragility functions of the loss-based limit states. Given the probability of being

in recovery-based limit state i, the probability of exceeding that limit state is taken as the sum of the probabilities of occurrence of all limit states equal to and greater than i.

$$P(RBDS > rbds_i \mid S_d) = \sum_{i}^{n_{rbds}} P\left(RBDS = rbds_i \mid S_d\right) \quad \textbf{(2-4)}$$

We can then use Equation 2-4 to compute the median spectral displacement, $\bar{S}_{d,rbds_i}$ and dispersion β_{rbds_i} for the recovery-based limit state fragilities. Figure 2-2, Figure 2-3, Figure 2-4 and Figure 2-5 provide a comparison of the recovery- and loss-based fragility functions. The parameters that define the recovery-based fragility functions for the wood light frame construction type, W1 (all code levels included) is summarized in Table 2-3.

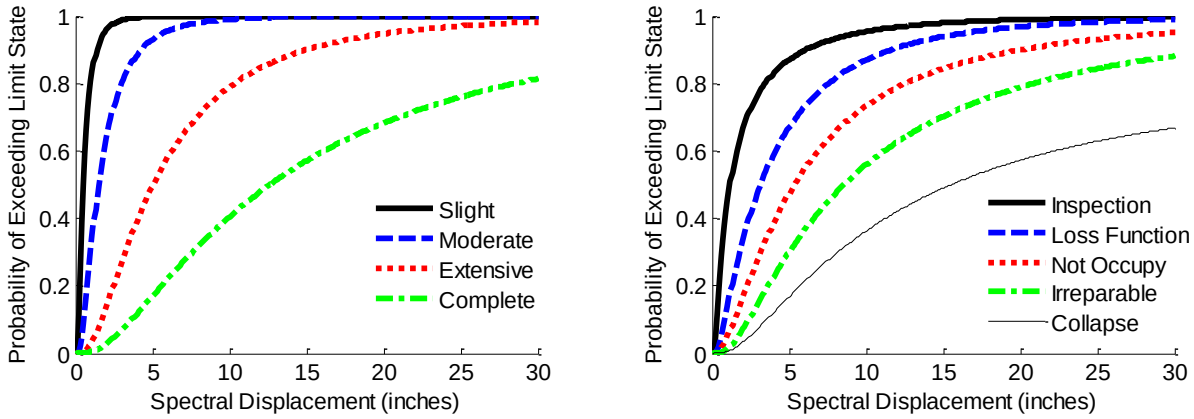


Figure 2-2. Fragility curves for (a) loss-based and (b) recovery based limit states for light wood frame buildings (W1) with high-code seismic design.

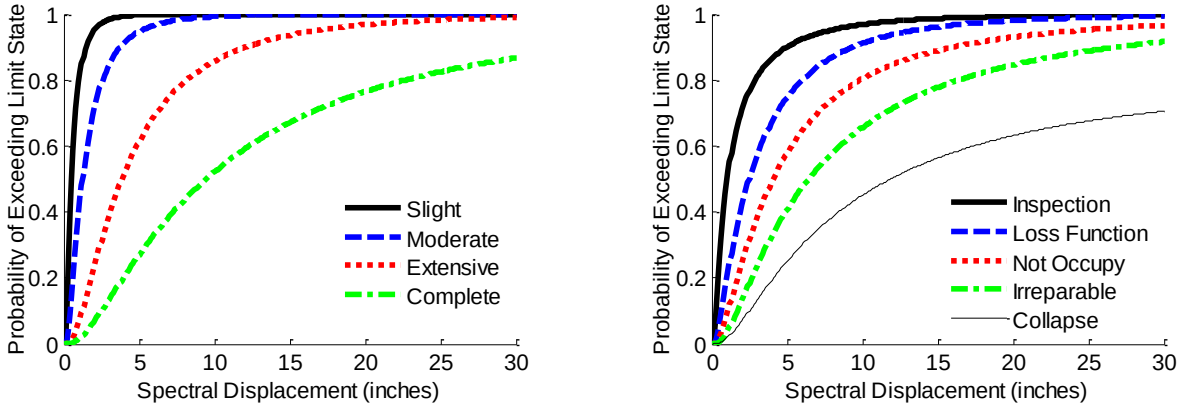
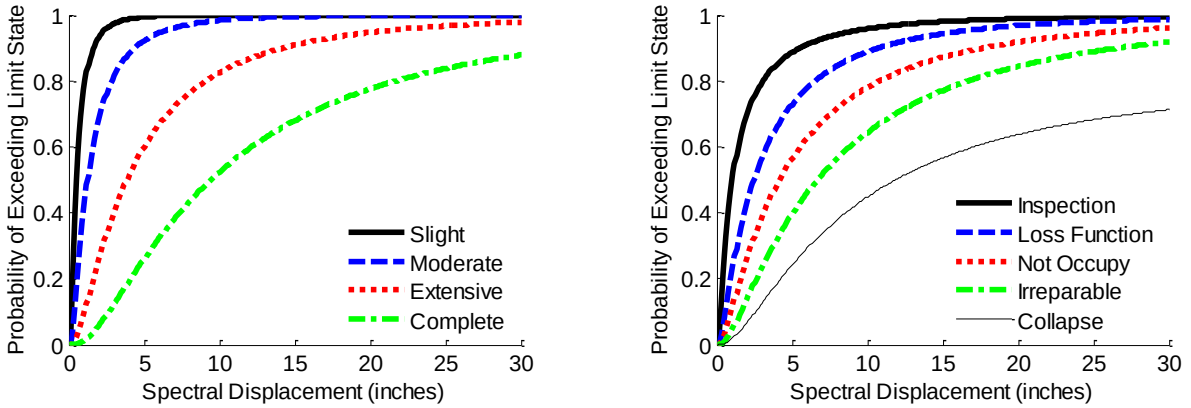
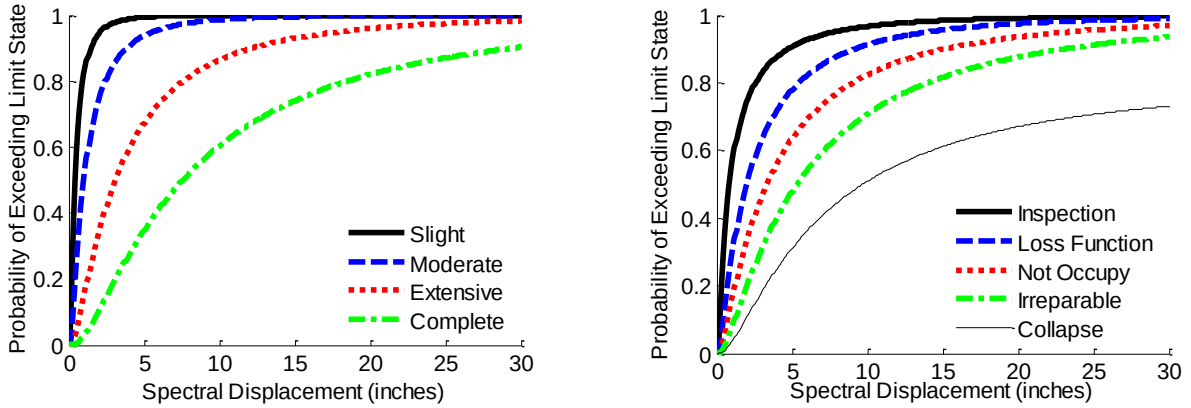


Figure 2-3. Fragility curves for (a) loss-based and (b) recovery based limit states for light wood frame buildings (W1) with moderate-code seismic design.



(a) (b)

Figure 2-4. Fragility curves for (a) loss-based and (b) recovery based limit states for light wood frame buildings (W1) with low-code seismic design.



(a) (b)

Figure 2-5. Fragility curves for (a) loss-based and (b) recovery based limit states for light wood frame buildings (W1) with pre-code seismic design.

Table 2-3. Fragility function parameters for recovery-based limit states for light wood frame buildings, W1: (a) Median spectral displacement at limit state exceedance and (b) log-standard deviation.

Code Level	Median Spectral Displacement for Exceeding Limit State				
	LS1 Inspection Triggered	*LS2* Loss of Functionality	*LS3* Unsafe to Occupy	*LS4* Damaged Beyond Repair	*LS5* Collapse
High-Code	1.2	3.1	5.4	8.5	15.6
Moderate-Code	1.1	2.5	4.1	6.4	11.9
Low-Code	1.0	2.5	4.0	6.1	11.9
Pre-Code	0.8	2.0	3.4	5.3	9.8

(a)

Code Level	Log-Standard Deviation of Spectral Displacement at Limit State Exceedance				
	LS1 Inspection Triggered	*LS2* Loss of Functionality	*LS3* Unsafe to Occupy	*LS4* Damaged Beyond Repair	*LS5* Collapse
High-Code	0.80	0.81	0.85	0.97	0.99
Moderate-Code	0.84	0.86	0.89	1.04	1.07
Low-Code	0.93	0.98	1.02	0.99	0.99
Pre-Code	1.01	1.05	1.07	1.06	1.08

(b)

Building-Level Recovery Model

Overview

In the previous section, it was noted that the recovery modelling methodology incorporates a set of building performance limit states that specifically inform community seismic resilience including (i) damage triggering inspection, (ii) occupiable damage with loss of functionality, (iii) unoccupiable damage, (iv) irreparable damage and (v) collapse. The link to post-earthquake recovery is established by defining a characteristic recovery path that is associated with each state and the level of functionality associated with each state. A building recovery function was computed accounting for the uncertainty in the occurrence of each recovery path and its associated limit state. The outcome is a probabilistic assessment of recovery of functionality at the building level for a given ground motion intensity. The overall methodology is based on the work by (Burton, Deierlein, Lallemant, & Lin, 2015).

Building Recovery Paths

Five distinct recovery paths were defined based on the limit states discussed previously. The recovery paths are described using discrete functioning states and the time spent in each state. The functioning states represent the changing condition of the building with respect to its ability to facilitate its intended operation. The functioning states for modelling the recovery of shelter-in-

place housing capacity include (1) the building is unsafe to occupy (*NOcc*), (2) the building is safe to occupy but unable to facilitate normal operations (*OccLoss*) and (3) the building is fully functional (*OccFull*). Note that these three states are specific to the shelter-in-place metric and would need to be re-defined for other measures of functionality. The key to defining the functioning states are that (1) they must be explicitly linked to the building level limit states described earlier and (2) each functioning state must be associated with a quantifiable measure of functionality.

The building level recovery path is conceptually shown in Figure 2-6. It is a step function that describes the time spent in each of the discrete functioning states. The recovery path (and recovery function discussed later) is assessed over a pre-defined period referred to as the control time, T_{LC}; and T_{NOcc}, $T_{OccLoss}$ and $T_{OccFull}$ that denotes the time spent in the *NOcc*, *OccLoss* and *OccFull* functioning states, respectively. It is important to note that the functioning states that comprise the recovery path for a given building depend on the limit state of that building immediately following the earthquake. For example, a building that is in limit state LS_1 will only experience the *NOcc* and *OccFull* functioning states. On the other hand, a building that is in limit state LS_2 or LS_3 will experience all three functioning states. This is illustrated later in the discussion of building recovery paths. The time spent in each functioning state will also vary depending on the level of damage. For example, a building that is in limit state LS_4, which must be demolished and rebuilt, will spend a significantly greater amount of time in the *NOcc* state than a building in limit state LS_3, which only requires repairs.

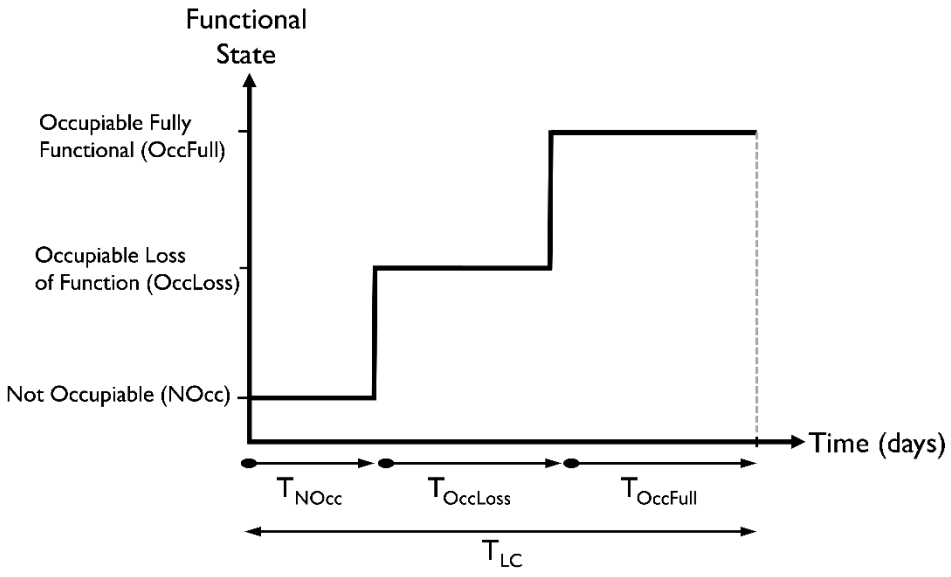


Figure 2-6. Conceptual illustration of recovery path for an individual building.

The recovery time for an individual building is defined as the period between the occurrence of the earthquake and the restoration of full functionality. The recovery time includes (1) the lead time which is the time required for building inspection and/or evaluation, finance planning, architectural/engineering consultations, a competitive bidding process and to mobilize for construction (Mitrani-Reiser, 2007), (2) the repair time needed to restore occupiability and (3) the repair time needed to restore functionality. The time needed to restore occupiability is taken as the time to complete repairs related to structural safety and internal access, whereas the time needed to restore functionality includes the additional time needed to repair/replace building systems, non-structural components and contents that are essential to the building functionality. Both the lead and repair times for structural and non-structural components depend on the limit state of the building immediately following the event. For example, a building that is in limit state LS_1 following an event (damage triggers inspection but the building is found to be safe to occupy and functional) will likely be green tagged and only be out of service for the time it takes to complete the inspection. On the other hand, a building that is in limit state LS_2 (building is safe to occupy but not functional) may receive a yellow tag, which would require detailed evaluations by a professional engineer prior to re-occupancy. A building that is red tagged (LS_3, LS_4 and LS_5) may require demolition or extensive repairs, triggering additional lead time for planning, architectural/engineering consultations, possible competitive bidding and mobilization for construction. Mitrani-Reiser, (2007) developed a performance-based approach to estimating repair times for both structural and non-structural damage, which incorporates the lead times for different tagging scenarios as well as the sequencing of repairs. In this study, Mitrani-Reiser's method is used to compute both the repair time needed to restore safety/accessibility and the repair time needed to restore functionality.

The recovery paths for each limit state were derived from the information provided in Table 2-4, which shows the relevant activities and time spent in each functioning state. The recovery paths are described as follows:

- *Recovery Path for LS_0*: This implies that the functionality of the building is not disrupted and the *OccFull* state is maintained throughout the period following the earthquake.
- *Recovery Path for LS_1*: This path is associated with the occurrence of LS_1 where the extent of damage triggers inspection but does not compromise the functionality of the building. It is comprised of the *NOcc* and *OccFull* states. The time spent in the *NOcc* state is the time

to complete inspections. Following inspections, the building is deemed occupiable and fully functional, immediately entering fully functional *OccFull* state.

- *Recovery Path for LS_2*: For LS_2, the recovery path includes all three functioning states. Like recovery path 1, the building initially enters the *NOcc* state until inspections are complete. Following inspections, the building enters the *OccLoss* state because, despite being safe to occupy, repairs will be needed to restore functionality. The time spent in the *OccLoss* state is determined by the repair time for those building systems, non-structural components and content that is essential to the building function. Completion of these repairs returns the building to the fully functional *OccFull* state.

- *Recovery Path for LS_3*: The recovery path for LS_3 also includes all three functioning states. Initially, the *NOcc* state includes the inspection and other lead times, along with the time to complete structural repairs needed to restore occupiability. Since LS_3 is associated with significant structural and non-structural damage, the lead time will include planning, design consultations, bidding and the time to mobilize for construction. Following the completion of structural repairs, the recovery will enter the *OccLoss* state during which the repairs needed to restore functionality are completed. The completion of these repairs would return the building to the *OccFull* state.

- *Recovery Path for LS_4*: In LS_4, where the building is irreparably damaged, the recovery path includes the *NOcc* and *OccFull* states, where the *NOcc* state includes the time to demolish and replace the damaged building. As the recovery of this building involves new construction, occupancy is not likely to be restored prior to full completion, which is why this path does not include the *OccLoss* phase.

- Recovery Path *for LS_5*: The recovery path associated with partial or complete collapse is very similar to that of the demolition case, the only difference being that LS_5 would not require any time to assess whether or not the building could or would be repaired. However, this additional time is likely to be insignificant compared to the time needed to replace the building, hence the recovery paths associated with LS_4 and LS_5 are essentially the same.

Table 2-4. Recovery path activities and times for each functioning state.

Recovery Path No.	Time/Acitivies in Functional State		
	NOcc	OccLoss	OccFull
0	0	0	T_{LC}
1	T_{INSP}	0	$T_{LC} - T_{INSP}$
2	T_{INSP}	T_{FUNC}	$T_{LC} - T_{INSP} - T_{FUNC}$
3	$T_{INSP} + T_{ASMT} + T_{MOB} + T_{OCC}$	T_{FUNC}	$T_{LC} - T_{INSP} - T_{ASMT} - T_{MOB} - T_{OCC} - T_{FUNC}$
4	$T_{ASMT} + T_{MOB} + T_{REP}$	0	$T_{LC} - T_{ASMT} - T_{MOB} - T_{REP}$
5	$T_{MOB} + T_{REP}$	0	$T_{LC} - T_{MOB} - T_{REP}$

T_{INSP} - Time to complete inspections
T_{FUNC} - Time to restore functionality
T_{ASMT} - Time to conduct engineering assessment
T_{MOB} - Time to moblize for construction
T_{OCC} - Time to complete repairs needed to restore occupiability/structural safety
T_{REP} - Time to replace building

Probabilistic Assessment of Recovery of Functionality at the Building-Level

Each functioning state can be linked to a quantifiable level of functionality. The functionality will typically be specified based on building owner/stakeholder and community resilience needs. For example, the functionality of a hospital might be measured by the number of available patient beds or patient waiting times for procedures offered by particular departments. The functionality associated with the *OccFull* state is equal to the pre-earthquake patient bed capacity and waiting time of the hospital. Obviously, the functionality associated with the *NOcc* state will be zero regardless of the measure of functionality since a building that is not occupiable will not be functional. The level of functionality assigned to the *OccLoss* state is less obvious and will vary based on the measure of functionality and the post-earthquake operating protocol for the facility. For example, in the case of a hospital facility that is in LS_2, where the hospital is occupiable but has lost some of its essential services, the hospital administration may choose to halt operations and close the facility until those services are restored. On the other hand, the emergency needs of the community may compel the administration to provide some reduced level of medical treatment that is possible with limited building services. In such cases, the *OccLoss* state can be assigned a level of functionality that is some fraction of the pre-earthquake capacity. Another example is the case of residential buildings in a community, where, from the perspective of the policy-makers, functionality is measured by housing capacity or number of persons housed. Where loss of certain building functions would not preclude short-term shelter-in-place requirements, the shelter-in-

place functionality could be determined by assuming the full pre-earthquake housing capacity is achieved for both the *OccFull* and *OccLoss* states. Alternatively, the expected housing capacity at the *OccLoss* state may need to account for the likelihood that the building is evacuated by its residents given the loss of a particular service, i.e.,

$$E[q(t) \mid OccLoss] = [1 - P(Evac \mid OccLoss)][q(t) \mid OccFull] \quad \textbf{(2-5)}$$

where $E[q(t) \mid OccLoss]$ is the expected housing capacity for a residential building in the *OccLoss* functioning state; $P(Evac \mid OccLoss)$ is the probability that the building is evacuated, given that it is safe to occupy but without some of its services; and $[q(t) \mid OccFull]$ is the housing capacity associated with the *OccFull* state or the pre-earthquake housing capacity. The $P(Evac \mid OccLoss)$ can be determined based on judgment informed by observations from past earthquakes. Knowing the level of functionality associated with each functioning state, the recovery paths for each limit state can be related to recovery functions, as illustrated in Figure 2-7 and calculated as follows:

$$[q(t) \mid LS_i] = \begin{cases} [q(t) \mid NOcc] & t < [T_{NOcc} \mid LS_i] \\ [q(t) \mid OccLoss] & [T_{NOcc} \mid LS_i] \le t < [T_{NOcc} + T_{OccLoss} \mid LS_i] \\ [q(t) \mid OccFull] & [T_{NOcc} + T_{OccLoss} \mid LS_i] \le t < T_{LC} \end{cases} \quad \textbf{(2-6)}$$

where $[q(t) \mid LS_i]$ is the time dependent building functionality given its immediate post-earthquake limit state LS_i; $[q(t) \mid NOcc]$, $[q(t) \mid OccLoss]$ and $[q(t) \mid OccFull]$ represents the level of functionality associated with the *NOcc*, *OccLoss* and *OccFull* states respectively; $[T_{NOcc} \mid LS_i]$ is the time from the earthquake to the end of the *NOcc* phase associated with limit state LS_i; $[T_{NOcc} + T_{OccLoss} \mid LS_i]$ is the time from the earthquake to the end of the *OccLoss* phase for limit state LS_i.

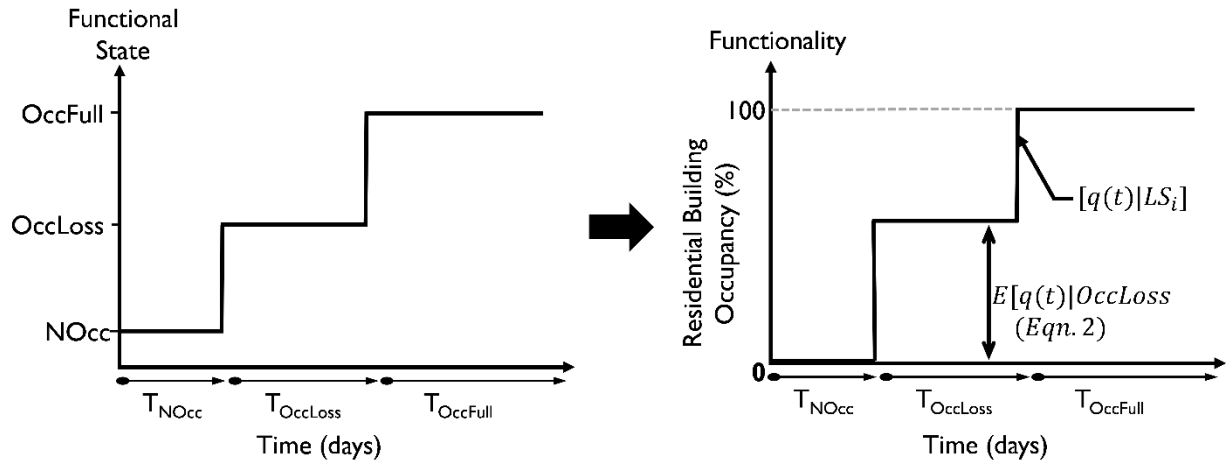


Figure 2-7. Conversion from recovery path to recovery function for residential building occupancy.

The building recovery function is computed accounting for the likelihood of the building being in each of the five limit states for a given ground shaking intensity. This is illustrated in the event tree, shown in Figure 2-8 where each limit state is associated with a unique recovery function, computed from Equation 2-7. Figure 2-8 also incorporates a sixth event that corresponds to damage below the threshold level that triggers inspection. The uncertainty in the building limit state and expected recovery is determined by the following

$$E[q(t) \mid IM] = \sum_{1}^{n_{ls}} [q(t) \mid LS_i] P[LS_i \mid IM] \tag{2-7}$$

where $E[q(t) \mid IM]$ is the expected recovery function given *IM* and $P(LS_i \mid IM)$ is the probability that the building is in the *ith* limit state for a given *IM* level.

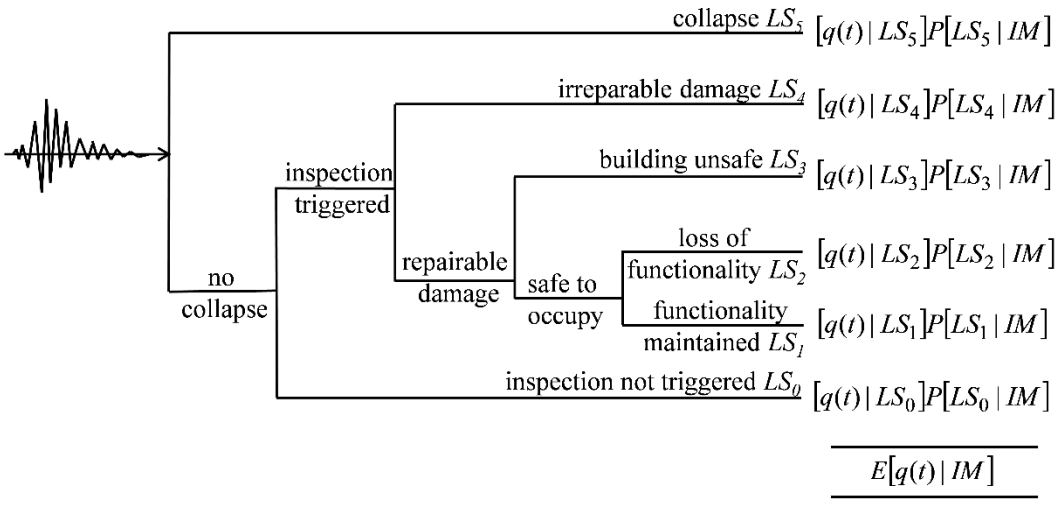


Figure 2-8. Limit state event tree used to assess building-level recovery.

Statistical Models for Linking the Rate of Recovery to Socio-Economic Variables

Overview

Externalities and socio-economic vulnerability will have a significant effect on the rate of recovery of communities. This section presents several statistical models in addition to logistic regression that could inform the relationship between the pace of recovery and some combination of predictor variables that are related to socio-economic factors. Three linear regression methods were utilized using the Napa recovery and resilience variables including (i) Ordinary Least Squares, (ii) Ridge, and (iii) The Least Absolute Shrinkage and Selector Operator. Among those methods, statistical significance is attained when a p-value is less than the significance level which indicates that the statistic is reliable, i.e. that predictor has a strong influence on recovery. Two machine learning methods were also incorporated: (i) Random Forest, and (ii) Support Vector Machine.

Ordinary Least Squares Regression (OLS)

OLS is one of the most basic and commonly used prediction techniques with applications in fields as diverse as statistics, finance, economics and engineering. It uses a linear combination of predictors to estimate the dependent variable which can be taken using the formula:

$$\boldsymbol{Y = X\beta + \varepsilon} \tag{2-8}$$

Y is an $n\times1$ vector of the dependent variable where n is the number of data points. X is an $n\times p$ matrix of the explanatory variables where p is the number of predictor variables. β is a $p\times1$ vector of the regression coefficients and ε is an $n\times1$ vector describing the random component of the linear relationship between X and Y.

The objective of *OLS* is to minimize the difference (residual) between the observed value of the dependent variable and predicted value by the linear approximation of the data, also called residuals.

$$\boldsymbol{\widehat{\beta} = argmin \sum_i (y_i - x_i^T \beta)^2 = (X^T X)^{-1} X^T Y} \tag{2-9}$$

To evaluate the statistical significance of the regression coefficient, the t-statistic is used, where implicit statistical hypotheses are $H_0: \beta = 0$ and $H_1: \beta \neq 0$. The t-statistic, which follows the t distribution with $(n-p)$ degrees of freedom, is computed as the ratio between the estimated regression coefficient and its standard error:

$$\boldsymbol{T = \frac{\widehat{\beta_J}}{se(\widehat{\beta_J})} \sim t(n-p)} \tag{2-10}$$

where $se(\hat{\beta}_j) = \sqrt{s^2(X^TX)^{-1}_{jj}}, s^2 = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-p}, \hat{\varepsilon} = \mathrm{Y} - \mathrm{X}\hat{\beta},$

The *p-value* is the probability of obtaining at least as extreme results given that null hypothesis is true, i.e. $p = Probability(t > |T|)$. If *p-value* is smaller than significance level α, the predictor attains statistical significance and therefore rejecting the null hypothesis. α is mostly often set at 0.05.

Ridge Regression

Ridge Regression is employed when analyzing multiple regression data that suffers from multicollinearity (Groetsch, 1984). Multicollinearity, which is the existence of strong correlations between predictor variables, can lead to inaccurate estimates of the regression coefficients, inflate the standard errors of the regression coefficients and deflate the partial t-tests for the regression coefficients.

Ridge regression penalizes the size of the regression coefficients by imposing a constraint on the sum of the squared values (the L_2 norm) of the predictor coefficients:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{argmin}\left[\sum_i (\boldsymbol{y}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 + \boldsymbol{\lambda}\sum_{j=1}^{p}\boldsymbol{\beta}_j^2\right] = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{\lambda}\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \quad \textbf{(2-11)}$$

To select λ, cross validation is conducted whereby, the data is split into K folds. The regression model is developed using $\mathrm{K} - 1$ folds and the test error is evaluated using the fold that was excluded from the model fitting. The best choice of λ would be the one that provides the least test error.

T-statistic in ridge is computed in a similar manner to OLS (Equation 2-10):

$$where\ se(\hat{\beta}_j) = \sqrt{s^2[(X^TX + \lambda I)^{-1}X^TX(X^TX + \lambda I)^{-1}]_{jj}}, s^2 = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-p}, \hat{\varepsilon} = \mathrm{Y} - \mathrm{X}\hat{\beta}$$

The Least Absolute Shrinkage and Selector Operator Regression (LASSO)

The *LASSO* regression is another method that addresses multicollinearity by doing both model selection and coefficient shrinkage (Tibshirani, 1996). Using L_1-penalization,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{argmin}\left[\sum_i (\boldsymbol{y}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 + \boldsymbol{\lambda}\sum_{j=1}^{p}|\boldsymbol{\beta}_j|\right] = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{\lambda}\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \quad \textbf{(2-12)}$$

LASSO is preferred over ridge regression when there is an assumption that the solution is sparse, i.e. many $\beta_i = 0$, because L_1 regularization shrinks some of the predictor coefficients to zero. If there

is no assumption of the sparse feature, ridge regression is usually preferred since ridge only shrinks the regression coefficients.

Random Forest

The random forest is a machine learning algorithm which combines the strategies of bagging and randomly selecting features to classify and regress data (Breiman, 2001). Specifically, random forest constructs a multitude of decision trees using training data and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. Bagging involves applying a majority vote (selecting the path with the greatest number of outcomes) for classification or prediction after many large trees are independently constructed using bootstrap resampled versions of the training data. Random forests change how the classification and regression trees are constructed. In normal trees, each node is split using the best split among all variables. In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node.

The algorithm of random forest is as follows:

1. Draw bootstrap samples from training dataset.
2. For each of sample tree, the Classification and Regression Tree (CART) is applied first. The idea of CART is to recursively divide the space into rectangular subspaces until satisfying some criteria of classification or regression.
3. In each tree, randomly selected features need to be incorporated to modify CART. The best split (based on the Gini index (Menze, et al., 2009) among a subset of predictors is chosen randomly rather than using all variables.
4. Classify or regress data by major vote on all individual tress.

Support Vector Machine (SVM)

SVM is another widely used supervised learning method that is used to analyze data and perform classification and regression. In *SVM* regression, the input *n*-dimension features *x* are first mapped onto a high *m*-dimensional space using some fixed mapping (Smola & Vapnik, 1997). In this study, the Gaussian Radial Basis mapping is used, then linear regression is used to construct the model in this space. The linear model in the *m*-dimension feature space could be defined as:

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\omega}) = \sum_{i=j}^{m} \boldsymbol{\omega}_j . \boldsymbol{g}_j(\boldsymbol{x}) + \boldsymbol{b} \qquad \textbf{(2-13)}$$

where, $g_i(x), i = 1, \dots, m$ denotes the transformation using Gaussian Radial Basis.

SVM regression uses the ε-insensitive loss function proposed by (Smola & Vapnik, 1997):

$$\boldsymbol{L}\big(\boldsymbol{y}, \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w})\big) = \mathbf{max}(\mathbf{0}, |\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w})| - \boldsymbol{\varepsilon}) \quad \textbf{(2-14)}$$

where ε is the tolerable bandwidth. At the same time, it also tries to reduce model complexity by minimizing $\|w\|^2$. Combining these two techniques, the following primal optimization problem is formed:

$$\boldsymbol{min}\left\{\tfrac{1}{2}\|\boldsymbol{w}\|^2 + \boldsymbol{C}\sum_{i=1}^{n}(\boldsymbol{\xi}_i + \boldsymbol{\xi}_i^*)\right\} \quad \textbf{(2-15)}$$

$$\text{s.t.} \begin{cases} y_j - f\big(x_j, w\big) - \varepsilon \le \xi_i \\ f\big(x_j, w\big) - y_j - \varepsilon \le \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \ i = 1, \dots n \end{cases}$$

The solution to the above dual problem is given by:

$$\boldsymbol{f}(\boldsymbol{x}) = \sum_{i=1}^{k}(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*)\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}) \quad \textbf{(2-16)}$$

$$\text{s.t.} \begin{cases} 0 \le \alpha_i \le C \\ 0 \le \alpha_i^* \le C \end{cases}$$

where, $K(x_i, x) = \sum_{j=1}^{m} g_j(x_i) g_j(x)$, and k is the number of support vectors.

Community-Scale Recovery Functions

Figure 2-9 shows an overview of the framework used to generate the community-level recovery functions. The performance-based earthquake engineering framework is applied to each building within the target community, incorporating the limit states described earlier; the outcome of which is a recovery function that is generated for individual buildings. The function describing community-level recovery is obtained by aggregating the recovery curves for the individual buildings after accounting for the variation and spatial correlation of shaking intensity at each site, the effect of externalities and other socio-economic factors. Note that the contribution of individual buildings to the functionality of the region depends on the type of building and measure of functionality. This aggregation of building-level functionality would require quantifying the contribution of each building to the defined measure of community function. The housing recovery function is described by the following equation:

$$\big[Q(t) \mid EQ_j\big] = \sum_{i=1}^{n_{bldg}} E\big[q_i(t) \mid IM_i, EQ_j\big] \quad \textbf{(2-17)}$$

where $E[Q(t) \mid EQ_j]$ describes community recovery for scenario earthquake *j*, $E[q_i(t) \mid IM_i, EQ_j]$ describes the expected recovery curve for building *i* at a given ground motion *IM* level resulting from scenario earthquake *j*, and n_{bldg} is the number of buildings in the community.

The long-term effects of an earthquake on a community can also be described by the cumulative loss of functionality over the course of the recovery period. For example, the loss of housing capacity over the recovery period measured in "person-days" can be computed from a community-level recovery curve that has number of residents housed by the community as the measure of functionality. This cumulative loss in functionality for a particular earthquake event is illustrated in Figure 2-9 and can be described by the following equation:

$$[LQ_{RE} \mid EQ_j] = \int_{T_E}^{T_{RE}} (Q_0 - Q(t))dt \qquad \textbf{(2-18)}$$

where $[LQ_{RE} \mid EQ_j]$ is the loss of functionality over the recovery period for scenario earthquake *j*, Q_0 is the pre-earthquake level of functionality, T_E is the time of the earthquake and T_{RE} is the time at full recovery.

Equation 2-18 describes the cumulative loss of functionality for a single scenario earthquake. Multiple scenario earthquakes can be considered and used to describe the annual exceedance rate for specified amounts of cumulative loss. This is obtained by computing the cumulative loss for multiple earthquake scenarios each with a different magnitude, location and annual rate of occurrence. The rate of exceedance, λ, for specified loss levels is estimated by summing the occurrence rate for all scenarios in which the loss threshold on interest is exceeded.

$$\lambda_{LQ_{RE} \geq lq} = \sum_{j=1}^{J} w_j I(LQ_{RE} \geq lq) \qquad \textbf{(2-19)}$$

where w_j is the occurrence rate for scenario j, and lq is the cumulative loss threshold. The indicator function $I(LQ_{RE} \geq lq)$ is set equal to 1 if the argument $LQ_{RE} \geq lq$ is true and 0 otherwise.

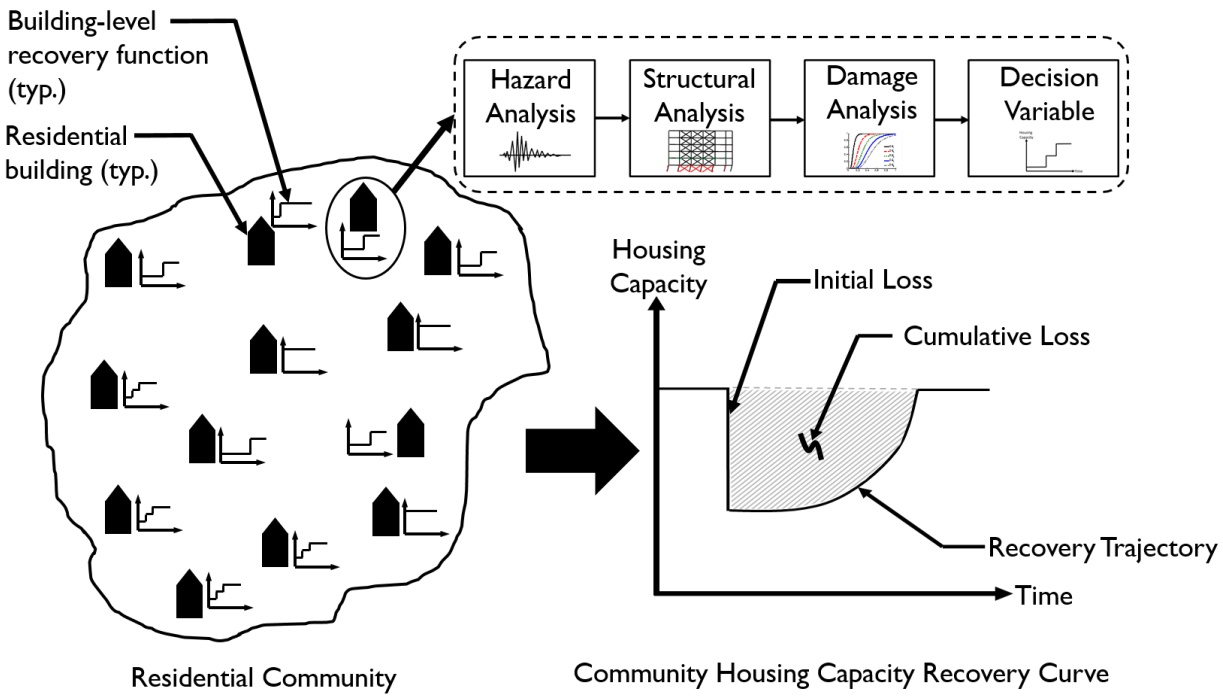


Figure 2-9. Conceptual illustration of recovery modelling framework.

3. Software Tool for Recovery Modelling

At the core of the recovery modelling project is an open-source computational tool that utilizes the OpenQuake-Engine (Silva, Crowley, Pagani, Monelli, & Pinho, 2014) and the recovery framework described herein to generate community scale recovery projections. Here, the methods, metrics, and recovery framework were incorporated into the existing OpenQuake Integrated Risk Modelling Toolkit, resulting in a dynamic and user-friendly software for generating both building-by-building and community level post-earthquake recovery predictions. The OpenQuake Integrated Risk Modelling Toolkit, developed by GEM, allows users to: 1) incorporate their local knowledge and data; 2) develop composite indicators (or indices) to measure social vulnerability and/or disaster resilience; 3) integrate these indices with physical risk estimates from OpenQuake or other software platforms; and 5) visualize the results. Following the latest enhancement of the software, it is now possible to utilize the OpenQuake-Engine to generate damage estimations, to generate building level and/or community level recovery functions based on the incorporated recovery framework, and to visualize and save the results within the QGIS environment. The software tool is transparent, and users can adjust the source code to their needs. The OpenQuake Integrated Risk Modelling Toolkit is available on the QGIS Plugin Repository.

In the following steps, a brief description of the basic workflow to develop an end-to-end recovery prediction is presented, where the main features and capabilities of the tool are highlighted. For the sake of demonstration, the recovery of a random sample of the residential buildings of the city of Napa is utilized as a use case.

STEP 1: Preparation of the input files to launch an OpenQuake-Engine analysis

The recovery modelling algorithm requires users to provide a CSV file containing the probability of exceedance of each limit state for each individual building in the exposure model. The latter can be computed by running a Scenario Damage Assessment, which is a type of analysis supported by the risk component of the OpenQuake-Engine. The input files necessary for running a scenario damage calculation and the resulting output files are depicted in Figure 3-1. For technical details, definitions and examples of each component, readers are referred to (Silva, Crowley, Pagani, Monelli, & Pinho, 2014).

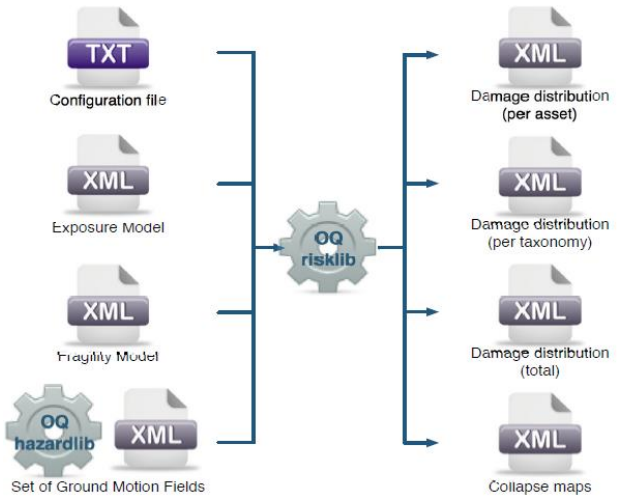


Figure 3-1. Scenario Damage Calculator input/output structure.

Figure 3-2 shows the window that requests users to upload the input files and run the scenario damage calculation.

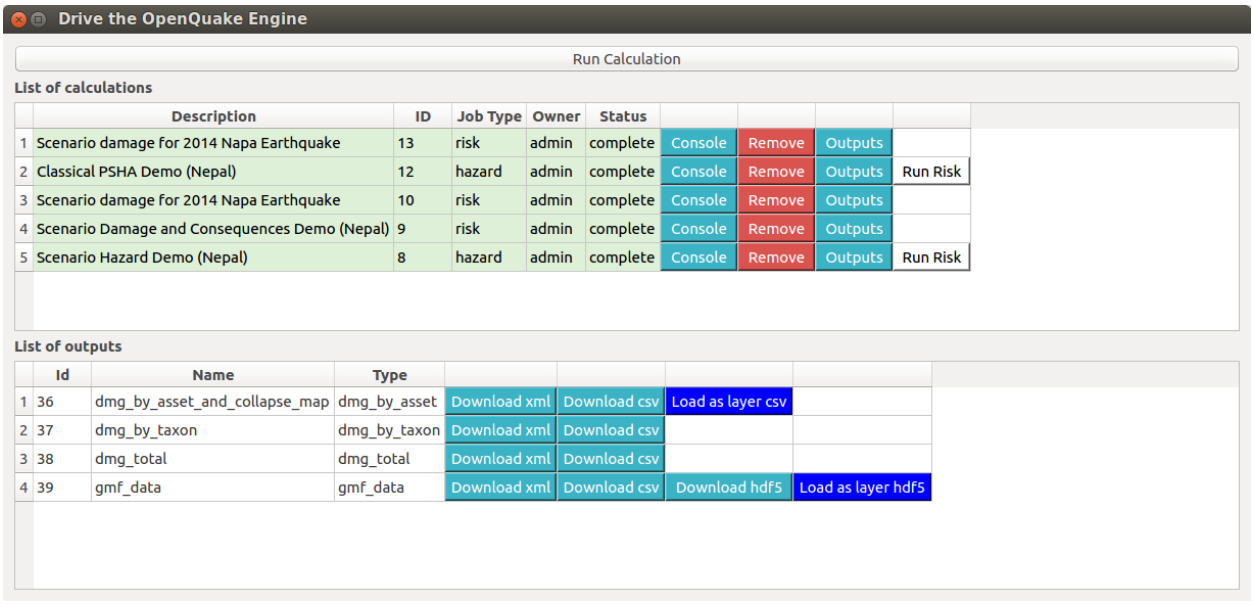


Figure 3-2. Pop-up window to run the OpenQuake Engine server.

It should be noted that the OpenQuake-Engine needs to be installed in the user's machine to run a calculation within the QGIS environment.

STEP 2: Preparation of the input files to run the recovery modelling algorithm

Table 3-1 presents the input files necessary to perform the recovery modelling analysis. The files should be adjusted to the available data and needs of the user.

Table 3-1. Required input files for the recovery modelling algorithm.

Input File	Short Explanation
Damage by asset (csv)	File that contains the mean probabilities of exceedance of each damage state for each individual building; output of the Scenario Damage Calculator of the OpenQuake-Engine
Assessment Times (txt)	Time to conduct engineering assessment
Inspection Times (txt)	Time to complete inspections
Mobilization Times (txt)	Time to mobilize for construction
Recovery Times (txt)	Period between the occurrence of the earthquake and the restoration of full functionality
Repair Times (txt)	Time to replace building
Repair Times Dispersion (txt)	Defines the level of uncertainty associated with the repair times
Lead Time Dispersion (txt)	Defines the level of uncertainty associated with the lead time
Transfer Probabilities (csv)	Discrete probability distribution of building level damage states (from OpenQuake)
Number of Damage Simulations (txt)	Number of damage realizations used in Monte Carlo Simulation

STEP 3: Conversion of the "Damage by asset" CSV file to a shapefile

As Figure 3-2 demonstrates, the list of the outputs from the Scenario Damage calculation can be visualized. The tool offers the possibility to load the "Damage by asset" CSV file as a QGIS vector layer, stored in the user's computer as a shapefile. In addition, it is possible to automatically style the layer with respect to a chosen damage state. Alternatively, the user can upload on QGIS the "Damage by asset" CSV file, structured in the same format as produced by the OpenQuake Engine, and save it as a shapefile.

At this point, the user may choose between two workflows on how to proceed to the generation of single buildings and/or community level recovery curves.

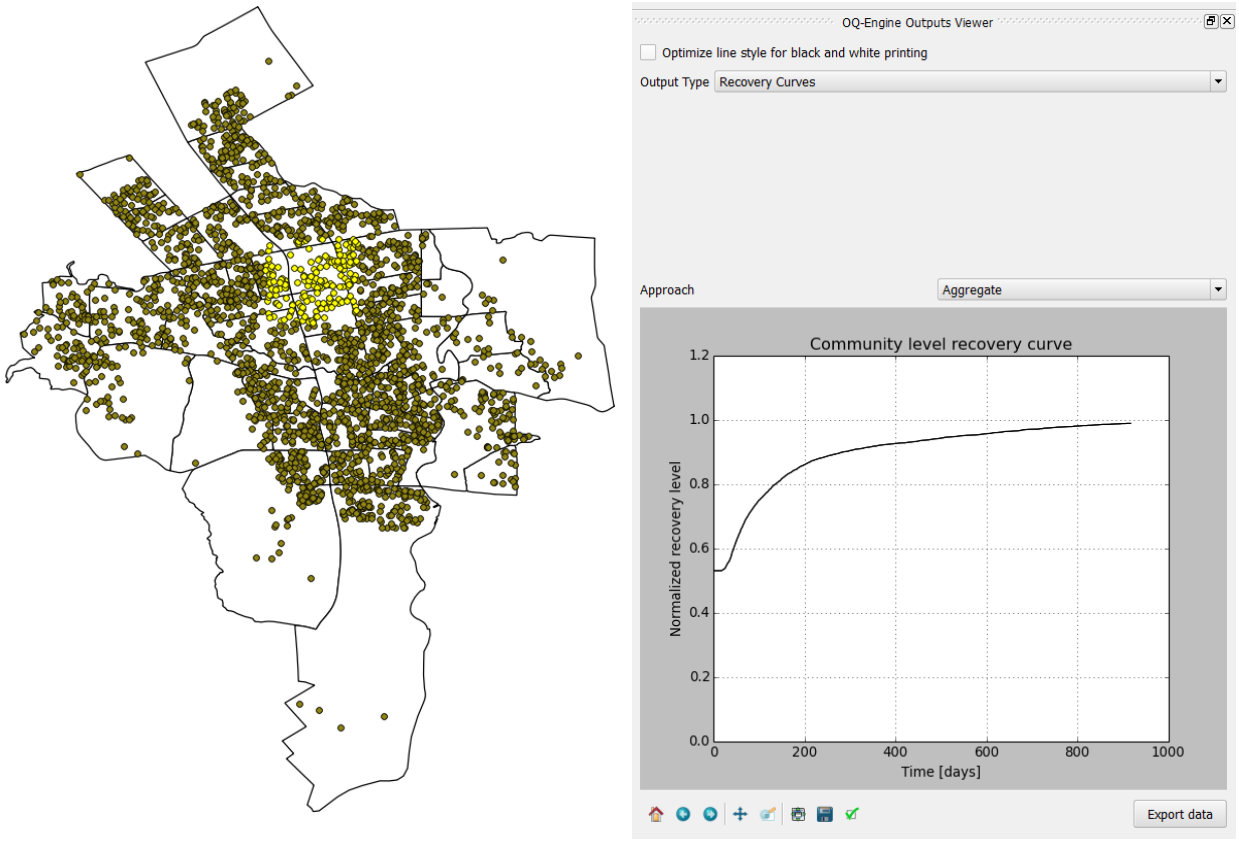
Workflow 1

The user can select individual buildings (or a group of buildings) and the respective recovery curve (single or aggregated) is automatically developed. The curve can be edited, digitized and exported as a CSV, as well as saved as an image. As shown Figure 3-3b, the user is required to select one of two available algorithmic approaches regarding the estimation of the recovery time (see Table 3-2) and, more importantly, to request the development of recovery curves by setting the Output Type

tab to "Recovery Curves". Figure 3-3 illustrates an example of the aggregated recovery curve of a set of selected buildings (highlighted with yellow).

Table 3-2. Short explanation of the aggregated and disaggregated approach for the estimation of the recovery time.

Approach for estimation of recovery time	Short Explanation
Aggregated approach	Building-level recovery model as a single process
Disaggregated approach	Building-level recovery modelled using four processes: inspection, assessment, mobilization and repair



a) b)

Figure 3-3. Aggregated recovery curve of a set of selected buildings (designated with yellow).

It should be emphasized that the integration of the recovery modelling algorithm in the QGIS software enables the users to adapt the workflow to their needs, leveraging all the features provided by the QGIS framework. The QGIS Processing Toolbox gives access to a wide variety of geoalgorithms, seamlessly integrating several different open-source resources, such as R, SAGA or GDAL. For instance, a SAGA algorithm, the "Add Polygon Attributes to Points", can be used to aggregate by zone a set of selected assets, resulting in relating each asset to the identifier of the

geographical area (zone) where it belongs. Following, the selection of the set of assets to be considered in the analysis can be performed in several different ways. The user can directly select points by clicking them on the map, or select points by using a formula. If points have been labeled with the identifier of the zone, the selection can be done with respect to the zone identification (or ID).

Workflow 2

Initially, as shown in Figure 3-4, the user must select the layer containing the information regarding damage state probabilities per asset (see **STEP 1**), after which a specific recovery time approach (Aggregate/Disaggregate) shall be opted. Here, it is possible to upload the layer of the study area with zonal geometries and generate aggregated recovery curves by zones. To exemplify, Figure 3-5 illustrates the block groups (zones) of the city of Napa, California and the aggregated recovery curve for the block group with the ID of 8032.

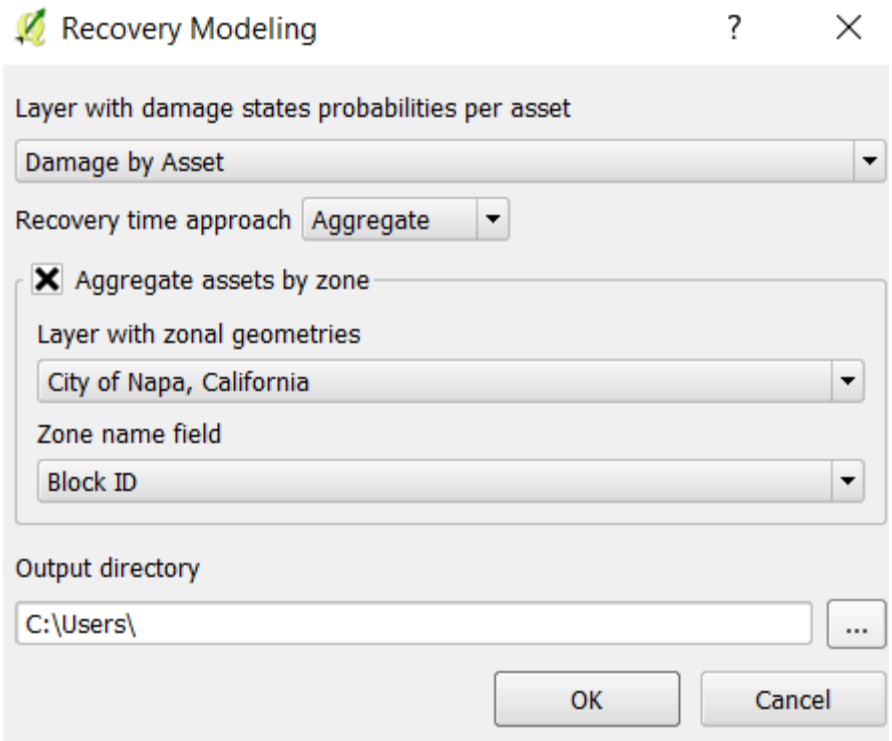


Figure 3-4. Pop-up window to run the recovery modelling algorithm.

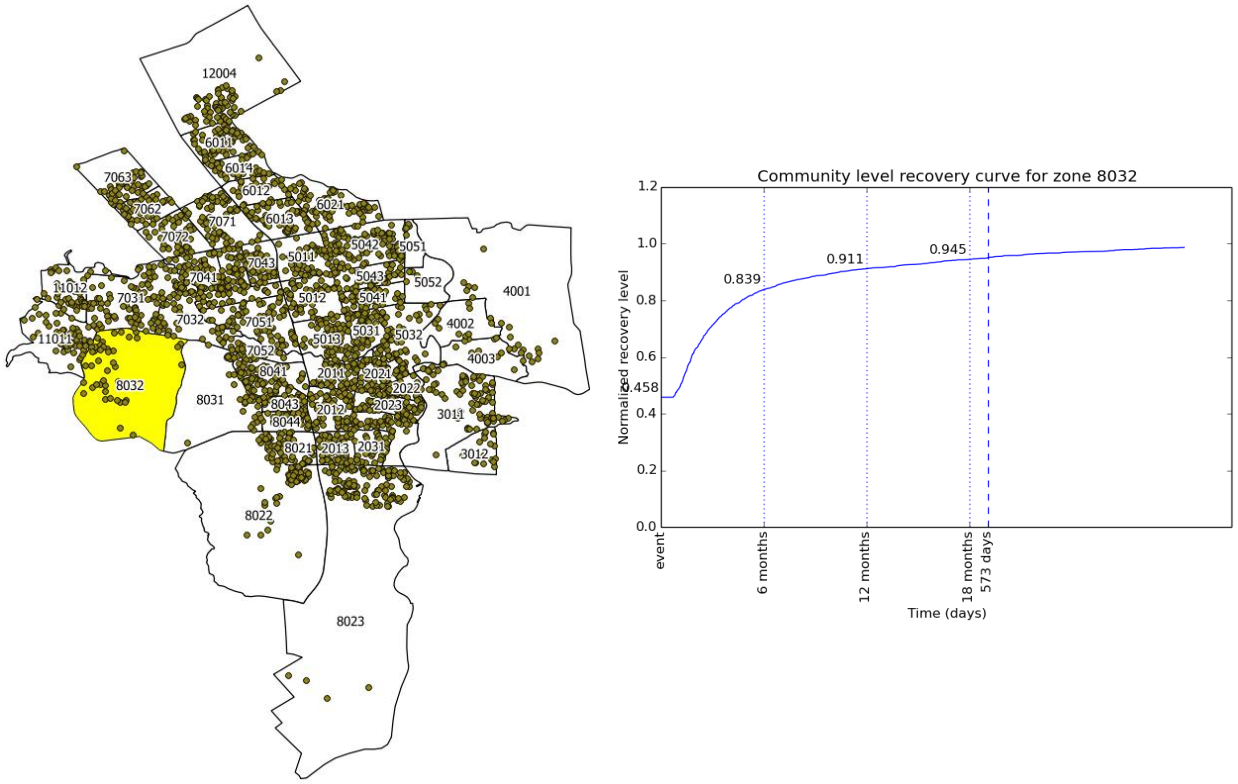


Figure 3-5. Community level recovery curve for the zone (block group) with an ID of 8032.

By unchecking the "Aggregate assets by zone" box (Figure 3-4) the algorithm generates a single community recovery curve by aggregating the recovery curves of all the buildings within the region. The graphs, similar to the one shown in Figure 3-5, are saved in the output directory designated by the user. In addition, building-by-building recovery curves are digitized and saved as text files (.txt) in the same output directory. The data can be further used (e.g. with Microsoft Excel) to generate and visualize individual building recovery curves that may be of interest to the user.

4. CONCLUSIONS

The road "back to normality" following a damaging earthquake is often rough and long. It is, therefore, of great value to monitor and study the recovery trajectory of a community, compile lessons learned, identify the factors that may impede the process and work towards their mitigation. Not surprisingly, the time of a community to recover is not only affected by the extent of the physical damage, but depends on many external factors strongly associated with the socio-economic conditions of the affected community. For example, the absence of earthquake insurance may burden homeowners with the rebuilding costs, which, especially in low-income families, may considerably hinder the recovery. Resilient communities are in a better position to absorb and withstand the impacts of a damaging earthquake and recover faster. A resilient community acknowledges its vulnerabilities, promotes measures to mitigate them and has pre-disaster plans in place to better and more efficiently respond and subsequently recover following a devastating earthquake. The work presented herein aims at facilitating stakeholders and decision makers to take more informed and efficient actions that primarily consider the needs of the most vulnerable groups of their communities.

Being state-of-the-art methodologies, it is acknowledged that further study and validation of the effectiveness of the proposed approaches is crucial. The methodology in section 1, for example, was developed using the city of Napa and the 2014 South Napa Earthquake, as a real-world case study. Therefore, it is expected that it will be applicable to communities with similar social characteristics and structure. However, future work shall focus on investigating the applicability and reliability of the methodology to different regions and different perils.

- **Scale of analysis**: The Napa case study was conducted at the U.S. Census Block Group level of analysis. However, it is important to consider to what extent changes in scale might lead to contradicting results. At minimum, research should be conducted to better understand the association between damage, resilience, and recovery potential at different scales.

- **Indicator selection**: The variable selection process for the development of composite indicators of resilience was subjective and based on secondary source data. The latter provides fertile ground for continued work that focuses on alternate assessment standards such as approaches that make use of resilience scorecards that are highly customizable and make use of primary source data.

References

Breiman, L. (2001). *Random Forests. Machine Learning* (Vol. 45).

Burton, C. (2015). A Validation of Metrics for Community Resilience to Natural Hazards and Disasters Using the Recovery from Hurricane Katrina as a Case Study. *Annals of the Association of American Geographers, 105*(1), 67-86.

Burton, H., Deierlein, G., Lallemant, D., & Lin, T. (2015). Framework for Incorporating Probabilistic Building Performance in the Assessment of Community Seismic Resilience. *J.Struct.Eng.* doi:10.1061/(ASCE)ST.1943-541X.0001321

Cutter, S., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. (2008). A Place-Based Model for Understanding Community Resilience to Natural Disasters. *Global Environmental Change, 18*(4), 598-606.

Cutter, S., Burton, C., & Emrich, C. (2010). Disaster Resilience Indicators for Benchmarking Baseline Conditions. *Journal of Homeland Security and Emergency Management, 7*(1), 1-22.

Galloway, B., & Ingham, J. (2015). The 2014 South Napa Earthquake and its relevance for New Zealand. *Bulletin of the New Zealand Society for Earthquake Engineering, 48*(1).

Groetsch, C. (1984). *The theory of Tikhonov regularization for Fredholm equations of the first kind* (Vol. 105). Pitman Advanced Publishing Program.

Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression, Second Edition.* United States of America: John Wiley & Sons, Inc.

Menze, B., Kelm, B., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics, 10*(1). doi:10.1186/1471-2105-10-213

Mitrani-Reiser, J. (2007). *An ounce of prevention: Probabilistic loss estimation for performance-based earthquake engineering.* Ph.D. dissertation, Department of Applied Mechanics, California Institute of Technology, Pasadena, CA.

Silva, V., Crowley, H., Pagani, M., Monelli, D., & Pinho, R. (2014). Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment. *Natural Hazards, 72*(3), 1409-1427.

Smola, A., & Vapnik, V. (1997). *Support vector regression machines. Advances in neural information processing systems.*

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Methodological), 58*(1), 267-288.

Wood, H., & Newmann, F. (1931). Modified Mercalli Intensity scale of 1931. *Bull. Seism. Soc. Am., 21*, 277-283.